The Intention-Outcome Asymmetry Effect: How incongruent intentions and outcomes influence judgments of responsibility and causality

Arunima Sarin, David A. Lagnado, and Paul W. Burgess

University College London, United Kingdom

Author Note

Arunima Sarin, Department of Psychology, Harvard University; David A. Lagnado, Cognitive, Perceptual, and Brain Sciences Department, University College London; Paul W. Burgess, Institute of Cognitive Neuroscience, University College London

Correspondence concerning this article should be addressed to Arunima Sarin, Department of Psychology, Harvard University; Cambridge, MA, 02138, United States of America. E-mail: asarin@g.harvard.edu

Abstract

Knowledge of intention and outcome are integral to making judgments of responsibility, blame, and causality. Yet, little is known about the effect of conflicting intentions and outcomes on these judgments. In a series of four experiments, we combine good and bad intentions with positive and negative outcomes, presenting these through everyday moral scenarios. Our results demonstrate an asymmetry in responsibility, causality, and blame judgments for the two incongruent conditions: well-intentioned agents are regarded more morally and causally responsible for negative outcomes than ill-intentioned agents are held for positive outcomes. This novel effect of an intention-outcome asymmetry identifies an unexplored aspect of moral judgments and is partially explained by extra inferences that participants make about the actions of the moral agent.

1. Introduction

Fundamental to successfully navigating our daily social interactions is the ability to identify causally and morally responsible agents. This ability is critical for explaining and predicting behaviour (Coffman, 2011; Heider, 1958; Young & Saxe, 2011). Causal and moral analyses are theoretically distinct. A person can cause an outcome without warranting blame, for instance when an infant accidently shoots someone, or be blamed for an outcome they didn't cause, such as the parents who failed to hide the gun from the infant (Lagnado & Channon, 2008). Yet, a substantial body of research suggests that causal and moral analyses are also intricately intertwined. The exact nature of the relation, however, is debated. Some evidence demonstrates a hierarchical relation between the two according to which causation is a necessary precondition for moral judgments (Heider, 1958; Darley & Schulz, 1990). In contrast, other findings show a bidirectional influence according to which not only do causal judgments influence moral judgments but moral judgments in turn influence perceptions of causation (Alicke, 1992; Hitchcock & Knobe, 2009; Knobe & Fraser, 2008; Kominsky, Philips, Gerstenberg, Lagnado, & Knobe, 2015). An important reason for the interaction between the two kinds of judgments is that they both rely on some common underlying components.

Previous research has shown that people's causal and moral attributions critically depend on their knowledge of an agent's intentions and their knowledge of the outcomes of the agent's actions (Alicke, 2000; Cushman, 2008; Guglielmo, 2015; Malle, Guglielmo, & Monroe, 2014; Young & Saxe, 2011). How do intention and outcome interrelate? It is well-documented that when they accord – that is when good intentions lead to good outcomes, or bad intentions lead to bad outcomes – the task of making causal and moral judgments is straightforward (Cushman, 2008; Gino, Moore, & Bazerman, 2009; Pizarro, Uhlmann, & Bloom, 2003). Yet, occasionally things are more complicated: intentions and outcomes can conflict. What happens to our moral and causal judgments in situations of conflict? Recent research has placed a considerable focus

on this question with the aim of disentangling the relative contributions of intentions and outcomes (Alicke, 2000; Baron & Ritov, 2004; Cushman, 2008; Cushman, Young, & Hauser, 2006; Pizarro, Uhlmann, & Bloom, 2003; Young & Saxe, 2011). In the most frequently adopted strategy, participants are asked to provide judgments of blame, punishment, and/or permissibility after reading about an agent who either unintentionally causes a harmful outcome (accidental harm) or has a harmful intention but fails to bring it to fruition (attempted harm) (Cushman, 2008; Cushman, Sheketoff, Wharton, & Carey, 2013; Young & Saxe, 2011). The focus in this line of research has been on the comparison of cases that study the impact of the presence or absence of something bad. Cases of accidental harm have a negative outcome but no negative intent, while cases of attempted harm have a negative intent but no negative outcome. Moral judgments made on a day-to-day basis, however, often involve situations that juxtapose positive and negative mental states and outcomes. The question of how we assign responsibility under situations of valence incongruity is rarely examined. The aim of this research paper is to shed light on this question. More specifically, the aim is to understand how we judge responsibility when intentions and outcomes mismatch.

1.1 The problem of mismatched intentions and outcomes

What does it mean for intentions to mismatch with outcomes? Imagine the following scenario: *Sandra works for a company that has an important meeting coming up with prospective clients. The company has scheduled a presentation for the potential clients with the aim of getting them to sign the contract they are offering. Sandra likes her work immensely and wants the company to succeed. She decides to make the presentation on her own. However, the clients hate the presentation and the company loses the contract*. How responsible do you hold Sandra for the outcome? To what extent would you say she caused the outcome?

Sandra's case highlights the tension between intention and outcome, when both factors are present. On the one hand, she gave the presentation that directly led to the loss of the contract.

On the other hand, her intention in giving the presentation was to benefit the company. In judging her responsibility and causality for the outcome, which one of these two weighs more heavily? Previous research on moral judgments offers two distinct perspectives.

1.2 Hierarchical Perspective

This perspective organizes deliberation about causal and intentional factors into a hierarchy. According to this approach, "judgments of moral responsibility presuppose those of causation" (Darley & Schulz, 1990). In other words, establishing a clear and direct causal link between an agent and the outcome is necessary before holding an agent responsible, and is sometimes sufficient in itself to warrant high degree of responsibility (Fincham & Jaspars, 1980; Heider, 1958; Shaver 1985). The claim stems from Heider's (1958) pioneering work on attribution theory that equates analysis of responsibility to climbing a staircase. The assessment of a causal link between an agent and the outcome is the first step of the staircase followed subsequently by assessments of intentionality, foreseeability, and justifiability (Darley & Shultz, 1990; Fincham & Roberts, 1985; Heider, 1958; Shaver, 1985; Shultz, Schleifer, & Altman, 1981; Weiner, 1995).

The hierarchical approach leads to two important predictions. First, since causal analysis precedes the analysis of intention, knowledge of an agent's intentions should not affect judgments of causation. In the previous example, Sandra's benevolent intention should not change her causal relation to the loss of the contract, and she should be held highly causal. Second, assuming that a causal link between Sandra and the loss of the contract is acknowledged, questions about Sandra's responsibility should incorporate knowledge of her intention. The loss of the contract is at odds with Sandra's intention, making it an unintended consequence. This should translate into a reduced rating of responsibility, when compared to a case where her intention is consistent with the outcome (Provencher & Fincham, 2000; Weiner, 1995).

1.3 Intentional Perspective

The second perspective places an overriding consideration on the knowledge of intention. Existing work shows that intentionally carried out actions are judged more responsible and causal compared to actions carried out accidentally or unintentionally (Cushman, 2008; Lagnado & Channon, 2008, Young & Saxe, 2011, Weiner, 1995). Robust evidence demonstrating the positive relation between presence of intention and degree of responsibility and causality comes from laboratory experiments (Lagnado & Channon, 2008: McClure, Hilton, and Sutton, 2007), clinical settings (Provencher & Fincham, 2000; Weiner, 1995), and even from the court of law where the act of killing another person culminates in punishment for murder or manslaughter depending on the perpetrator's intent (Pillsbury, 2000).

This approach to judging responsibility and causality makes some distinctive predictions. First, since the presence or absence of intention is a pivotal factor in deciding responsibility, the approach would predict reduced responsibility judgments for unintentional consequences. In other words, we would expect lowered responsibility for Sandra for the loss of the contract as the loss was unintended. In this regard, the intentional account's prediction parallels that of the hierarchical account. However, a difference between the two accounts lies in their supposition of the mechanisms supporting this prediction. On the hierarchical account, the presence or absence of intentionality is factored in only after the establishment of a causal link between the agent and the outcome. The influence of intention and outcome on the overall judgment of responsibility is thereby unidirectional and hierarchical. In contrast, the intentional account is not wedded to a specific idea of directionality of influence. Therefore it could either be that the reduction in responsibility, due to unintentionality of the action, happens after a causal link has been established or that the assessment of intentionality itself affects the assessment of causality (Alicke, 1992; 2000; Lombrozo, 2010; Philips & Shaw, 2014). This brings us to the second set of predictions regarding causality ratings, one arising from each of the two possibilities. If the

relationship between intentions and responsibility judgments is conditioned upon the presence of causality, then we would expect high causal rating for Sandra, just like we do on the hierarchical account. However, if instead a bidirectional relationship exists between intention and causal assessment, such that knowledge of intention influences perceived causality (Alicke, 1992; 2000; Philips & Shaw, 2014), we would expect to see reduced causal ratings for Sandra.

1.4 Intention & Outcome

To summarize, both accounts – hierarchical and intentional – predict reduced ratings of responsibility for Sandra when her intention clashes with the outcome compared to a case where her intention is consistent with the outcome. Further, on the hierarchical account we expect a high degree of causal association between Sandra and the outcome, while on the intentional account we expect a reduced degree of causal association. Sandra's case and the accompanying predictions sketch a partial outline for the interplay between intention and outcome, when good intentions lead to negative outcomes. To complete the picture we need to include predictions for the opposite case where bad intentions lead to positive outcomes. Thus consider a different agent, Alesandra, who dislikes her work and makes the presentation with the intention to lose the contract. Despite this intention, the clients love her presentation and the company wins the contract. On the hierarchical account we would expect Alesandra to be held highly causal since she made the presentation that won the contract, but not highly responsible, as the win was unintended. On the intentional account we would not expect Alesandra to be held highly causal or highly responsible.

Before we put these predictions to test, it is important to highlight facets of this research that distinguish it from previous work. The predictions we have derived for the two cases come from research that has largely compared accidental harms with attempted harms. In other words, the point of focus have been situations in which neutral or absent intentions have led to harmful outcomes, and situations in which harmful intentions have led to neutral or status-quo outcomes

respectively. At a cursory glance, the two situations may seem to parallel the cases of Sandra and Alesandra respectively. However, a closer examination reveals difficulties with the analogy. Sandra has an intention that is opposed to her outcome, but the intention is not neutral or absent. It is present and made explicit. With this in mind, does her case mimic accidental harm or attempted benevolence? Similarly, though Alesandra's case could be construed as attempted harm, unlike prototypical cases, the outcome is not neutral. The outcome is the winning of the contract which is positive and opposed to her intention. Does Alesandra's case represent attempted harm or accidental benefit? How the two cases of Sandra and Alesandra are construed raises an important theoretical question about how blame and praise interact. Sandra and Alesandra's cases demonstrate important but previously unexamined cases of moral and causal judgments.

A final point to highlight is that most previous scenarios have used highly adverse outcomes, such as those in which one agent grievously injures or kills another (Shultz, Schleifer, & Altman, 1981; Gino, Moore, & Bazerman, 2009). While valuable for the understanding of some aspects of moral and causal decision-making, the situations lack ecological validity (Bauman, McGraw, Bartels, & Warren, 2014), making it hard to ascertain how moral and causal decisions are made by an average person during the course of his or her daily life.

To understand how moral and causal judgments are *typically* made on a daily basis, the present experiments employ familiar and quotidian scenarios including both positive and negative mental states and outcomes. Our aim is to understand how an incongruence between intention and outcome manifests itself in moral and causal judgments. The first experiment explicitly tests the predictions outlined by the hierarchical and intentional accounts. The second and third experiments focus on characterizing the extent to which the results from the first experiment generalize. Finally, the fourth experiment explores one potential explanation underlying the findings from the first three experiments.

2. Experiment 1

The first experiment explored how an incongruence between intention and outcome affected judgments of responsibility. Based on the predictions of the hierarchical and the intentional accounts it was hypothesised that responsibility judgments would be reduced for the two incongruent cases (good intentions – bad outcomes, bad intentions – good outcomes) when compared with the two congruent cases (good intentions – good outcomes, bad intentions – bad outcomes). Intention and outcome were thus varied on two levels: positive and negative. All conditions were presented to the participants through scenarios reflective of everyday situations such as making a presentation for a company or planning a family gathering. The dependent variable of interest was the degree to which agents in each condition were held responsible for the outcome.

2.1 Method

2.1.1 Participants

152 people participated in the experiment. 18 people were excluded for: leaving parts of the study incomplete (n = 12), incorrectly answering check questions (n = 3), and taking more than three times the average time to finish the study (n = 3). Of the remaining 134 participants, 59 (44%) were female. The ages of the participants ranged from 18 to 60 inclusive, with an average age of 27.63 (SD = 8.61). Participants were paid £1 to participate.

2.1.2 Design and Materials

Subjects were presented with 16 vignettes that manipulated knowledge of intention and outcome through four scenarios, making it a 2 x 2 x 4 within-subject design. An agent's intention and outcome were clearly stated in each scenario. Order of presentation was randomised for all participants. For a full list of all 16 vignettes, please refer to Appendix A. An example of the parametric variations of intention and outcome for the company scenario is shown in Figure 1.

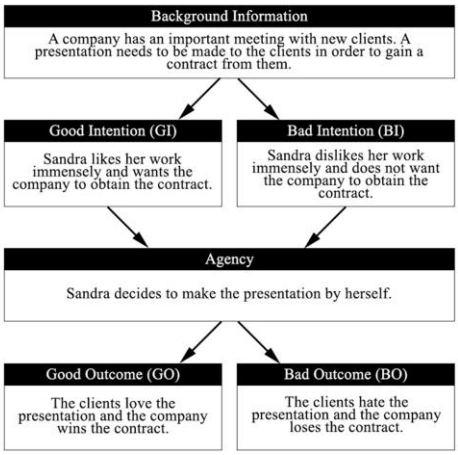

Figure 1. All combinations of intention and outcome for the company scenario.

The primary dependent variable of interest was the rating of responsibility. Following the presentation of a vignette, participants were asked to judge the responsibility of an agent for the outcome ("To what degree is Sandra's presentation responsible for the company winning the contract?"). The rating was obtained on a discrete slider ranging from 0 ("Not at all responsible") to 10 ("Completely responsible"). In addition, participants were asked a factual question to gauge their attention to the scenarios presented ("What did Sandra make for the company?"). Any participant failing to answer these questions correctly was dropped from analysis and not compensated. All other participants were compensated monetarily. Across all experiments, no participant was repeated.

2.2 Results

2.2.1 Effect of Scenario

A repeated measures ANOVA revealed a significant effect of scenario on ratings of responsibility $F(3,399) = 13.21$, $p < .001$, $\eta_p^2 = .090$. Post-hoc tests with Bonferroni correction

showed a statistically significant difference between the company scenario and the other three, even though the actual largest magnitude of difference was only 0.62 responsibility points. Consequently, scenario was included as a within-subjects factor in subsequent analyses.

2.2.2 Ratings of Responsibility

Mean ratings of responsibility for the four experimental condition were derived. Ratings were the highest for the bad intention-bad outcome condition (M = 8.64) followed by the conditions good intention-good outcome (M = 8.48), good intention-bad outcome (M = 7.06), and bad intention-good outcome (M = 5.42) respectively (see Figure 2).

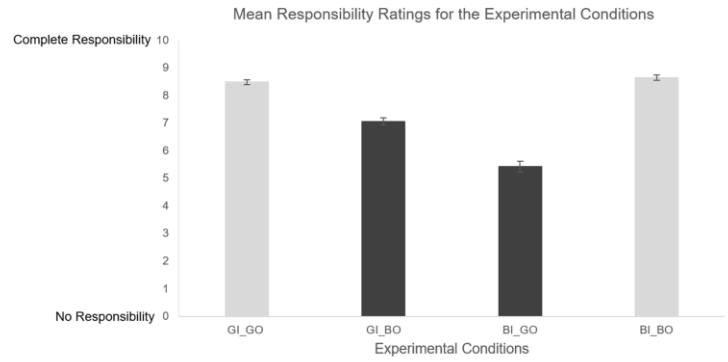

Figure 2. Mean ratings of responsibility for the four experimental conditions. Higher scores represent greater responsibility ratings. Error bars represent standard error of mean. Abbreviations: G = Good; B = Bad; I = Intention; O = Outcome

A three way repeated measures ANOVA revealed main effects for all three factors: intention, outcome, and scenario. For the main effect of intention $F(1,133) = 74.89$, $p < .001$, $\eta_p^2 = .360$, good intentions (M = 7.77) got a higher degree of responsibility than bad intentions (M = 7.03), $p < .001$. A main effect of outcome $F(1,133) = 101.12$, $p < .001$, $\eta_p^2 = .432$, showed that negative outcomes (M = 7.85) garnered higher responsibility ratings than positive outcomes (M = 6.95), $p < .001$. For the main effect of scenario $F(3, 399) = 13.21$, $p < .001$, $\eta_p^2 = .090$,

ratings of responsibility were the highest for the company scenario (M = 7.79) followed by those for the family gathering (M = 7.36), the restaurant (M = 7.30), and the theatre production (M = 7.16).

Significant two way interactions were observed for intention and outcome, $F(1, 133) = 324.89$, $p < .001$, $\eta_p^2 = .710$, intention and scenario, $F(3,399) = 6.34$, $p < .001$, $\eta_p^2 = .045$, and outcome and scenario, $F(3,399) = 4.90$, $p < 0.05$, $\eta_p^2 = .036$. There was also a significant three way interaction between intention, outcome, and scenario, $F(3, 399) = 8.24$, $p < .001$, $\eta_p^2 = .058$.

The main effects of intention and outcome and their interaction on the ratings of responsibility were checked for each of the four scenarios. A two way repeated measures ANOVA showed that all three of the effects were preserved within each of the scenarios.

2.3 Discussion

The results of Experiment 1 demonstrated a surprising interaction between intention and outcome in making responsibility judgments for the incongruent cases. Overall, the incongruent conditions received lower ratings than the congruent ratings. However, there was a novel interaction between intention and outcome for the incongruent cases. In scenarios involving good intentions and bad outcomes, like Sandra's, the agents were judged more responsible than in scenarios involving bad intentions and good outcomes, like those of Alesandra's. This asymmetry in the evaluation of the two incongruent conditions is surprising as neither the hierarchical nor the intentional perspective accounts for it.

On the hierarchical perspective the absence of a causal link between action and outcome, or the lack of desire to obtain a certain outcome would lower an agent's responsibility. The intentional account makes a similar prediction of reduced responsibility contingent upon the absence of intention with respect to the obtained outcome. Together, both accounts point to

reduced responsibility judgments for consequences that are unintended. Since neither account specifically discriminates between consequences differing in their valences, it can be argued that reductions in responsibility will be the same for both types of incongruence (good intentions-bad outcomes, bad intentions-good outcomes). Contrary to this prediction, participants treated the two cases differently. While responsibility ratings were reduced for both, the reduction was significantly greater for cases of a bad intentioned agent causing a positive outcome than cases of a good intentioned agent causing a negative outcome. In other words, good intentioned agents causing bad outcomes were held more responsible than bad intentioned agents causing a good outcome.

What might be the reasons for this asymmetrical evaluation? Prior to addressing this question, it is important to establish the robustness of this novel effect. Experiment 1 uses vignettes to present information to the participants. Since the type of scenario had a significant effect on the results, we first need to verify that the results hold under different scenarios. A second issue is the use of responsibility judgments as the dependent measure. Prior work in attribution research has criticised the word 'responsibility' for being polysemous. A question about responsibility could therefore be construed as a question about causality, blame, or even punishment (Fincham & Jaspers, 1980; Gerstenberg, Lagnado, & Kareev, 2010). Although related, each of these concepts are distinct. We need to establish whether the results obtained in Experiment 1 are specific to one of these judgments or generalize over the different dependent measures. The next two experiments address these two issues sequentially.

3. Experiment 2

This experiment sought to replicate the novel asymmetry for incongruent conditions using a different range of scenarios. In addition we tested for an effect of severity of outcome, which is often a factor in determining people's responsibility judgments.

Previous work on attribution of responsibility suggests that people rely on knowledge of the severity of an outcome, in addition to its valence, in making moral judgments (Medway & Lowe, 1975; Walstser, 1966; Shaver, 1970). Some studies contend that more severe outcomes garner harsher judgments (Medway & Lowe, 1975; Phares & Wilson, 1972; Shaw & Skolnick, 1971). DeJoy and Klippel (1984) presented participants with vignettes describing alcohol-related near-miss accidents and varied the level of unsafe behaviour as well as the severity of the accident. They found that regardless of the presence of unsafe behaviour, responsibility was assigned based on outcome information with more severe outcomes getting higher scores. Further, in the absence of outcome information, participants did not view very unsafe behaviour as significantly different from safe behaviour. However, other studies find no evidence for any impact of outcome severity on moral judgments (Arkkelin, Oakley, & Mynatt, 1979; Thomas & Parpal, 1987; Walster, 1967). Yet other studies show that an increase in outcome severity actually reduces degree of responsibility and blame (McMartin & Shaw, 1977; Shaw & McMartin, 1977). For instance, Shaw and McMartin (1977) presented participants with vignettes in which an agent caused a mild or severe accident. They found that with an increase in the severity of an outcome, responsibility attribution to an agent decreased. While evidence concerning the direction of impact of outcome severity on moral judgments is inconsistent, the use of outcome information in making moral judgments, appears robust (Mazzocco, Alicke, & Davis, 2004). Experiment 2 therefore systematically varies outcome information on two different levels of severity to examine its impact on responsibility judgments for the two incongruent cases. Since the focus is on examining the generalisability of the finding with respect to the stimuli, we continue to use responsibility ratings as the dependent measure in this experiment.

3.1 Methods

3.1.1 Participants

10 participants (n = 6, incomplete study; n = 3, failure to answer check questions; n = 1 more than three times the average time to complete the study) were eliminated from an initial sample of 114. The remaining 104 participants included 37 females (35.6%). All participants were between the ages of 18 and 57 (inclusive) with an average age of 28.73 (SD = 9.41). Participants were compensated with £0.92 for their participation.

3.1.2 Design and Materials

Participants were presented with 24 unique vignettes that arose from a 2 x 2 x 2 x 3 design with intention (good, bad), outcome valence (positive, negative), outcome severity (low, high), and scenarios making up the respective within-subject factors. The scenarios comprised of a gardening situation, a prom party, and a house redecoration. Like the scenarios of the first experiment, the present scenarios were chosen for their similarity to everyday life. For a full list of all 24 vignettes refer to Appendix B.

Participants provided ratings of responsibility using a slider identical to the one used in the first experiment. They also answered a factual question for each vignette.

3.2 Results

A four way repeated measures ANOVA was performed using intention, outcome valence, outcome severity, and scenario as the within-subject factors. Main effects were found only for intention $F(1,102) = 20.38$, $p < .001$, $\eta_p^2 = .167$ and scenario $F(2, 204) = 27.32$, $p < .001$, $\eta_p^2 = .211$. The severity of the outcome did not have any significant main effect $F(1, 102) = 2.93$, $p = 0.09$, ns. Averaging over the two severity conditions did however produce ceiling effect which was reflected in a smaller although still significant difference between the two incongruent conditions. Overall, good intentions got slightly higher ratings of responsibility (M = 7.90) than bad intentions (M = 7.53). Significant interaction effects were found for intention

and outcome $F(1,102) = 79.54$, $p < .001$, $\eta_p^2 = .438$, outcome and scenario $F(2, 204) = 5.56$, $p < 0.05$, $\eta_p^2 = .052$, and severity and scenario $F(2,204) = 19.00$, $p < .001$, $\eta_p^2 = .157$.

Since scenario had a significant main effect, the main effect of intention and the interaction between intention and outcome were checked individually for each of the three scenarios. The results were consistent with those obtained in the four way ANOVA (see Figure 3).

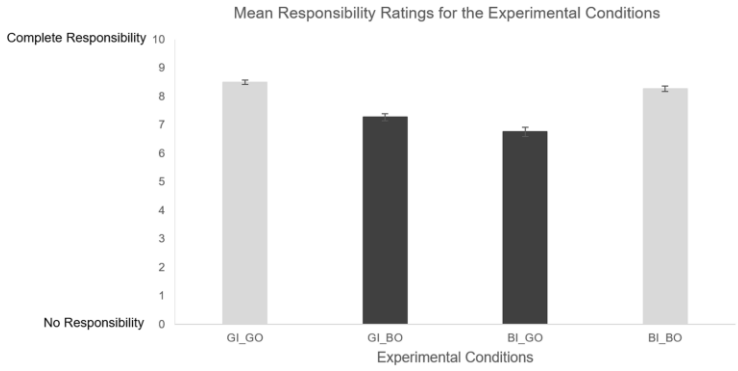

Figure 3. Mean ratings of responsibility averaged across outcome severity. Error bars represent standard error of the mean. Higher scores imply greater responsibility ratings. Abbreviations: G = Good; B = Bad; I = Intention; O = Outcome.

3.3 Discussion

This experiment had two aims. First, to assess the generalizability of the asymmetry observed in Experiment 1. The initial finding was replicated: overall, the two incongruent conditions received reduced ratings of responsibility compared to the congruent conditions, and most importantly, agents with good intentions - bad outcomes received higher responsibility than agents with bad intentions – good outcomes. This suggests that the observed asymmetry is not due to the specific scenarios used.

The second aim was to test the impact of outcome severity on judgments of responsibility. Outcome severity did not seem to affect the overall judgments of responsibility in our experiment. However, a closer inspection of the severity data revealed an interesting pattern. A significant interaction was reported for severity and scenario. With an increase in outcome severity, ratings of responsibility increased for the gardening condition, decreased for the prom party condition, and did not significantly change for the house redecoration condition. This pattern of responsibility judgments is complex, but existing research echoes similarly mixed findings.

According to Walster's (1966) defensive attribution hypothesis, ratings of responsibility increase with an increase in the severity of the outcome. However, according to Shaver's (1970) relevance hypothesis, it is the degree of situational and personal relevance felt by a participant that mediates the relationship between outcome severity and judgments of responsibility. The degree to which a situation seems relatable to a participant construes the situational relevance while the degree to which participants personally identify with a situation construes personal relevance. Shaw and McMartin (1977) found that high situational and high personal relevance produced a pattern of judgment predicted by the relevance hypothesis, scenarios of only high situational relevance led to an attribution pattern suggested by the defensive attribution hypothesis, and a lack of situational relevance eliminated the effect of outcome severity on judgments of responsibility all together.

A similar effect might be taking place in the present experiment. However, it is also possible that the impact of severity information is intrinsically related to the scenarios such that a severe outcome for one scenario may not be equivalently severe for another scenario. Some research has found support for a multidimensional aspect of outcome severity such that different dimensions (e.g. duration, mental/physical) may have different effects on ratings of responsibility (Wissler, Evans, Hart, Morry & Saks, 1997; Slain, Penrod, Garbin, & Stolle,

1998). Future research would be required to tease apart these factors systematically to better understand the relation between them.

4. Experiment 3

Experiment 3 assessed if the 'responsibility' response format used in the previous two experiments influenced the observed pattern of data. Previous research indicates that the word 'responsibility' could denote different meanings such as cause or blame (Fincham & Jaspers, 1980; Gerstenberg et al., 2010). To systematically test for this, Experiment 3 asked participants to assess the degree to which an agent's action was the *cause* of a particular outcome. In addition, participants were asked to assign blame or praise to agents. Since the type of scenario did not account for the observed asymmetry, the experiment used three of the original four scenarios from Experiment 1. The decision to employ three instead of four scenarios was motivated by the desire to make the duration of the experiment shorter. Since responsibility judgments were similar for each of the scenarios, one of the scenarios was picked at random and dropped. All participants received the same three scenarios.

4.1 Method

4.1.1 Participants

48 people took part in the experiment. After eliminating those who did not complete the study (n = 5) and those who failed to answer the check questions (n = 1), the remaining 42 participants were in the age range of 18 - 60 (inclusive) with an average age of 28.76 (SD = 9.25). 28 (66.7%) participants were males.

4.1.2 Design and Materials

Each participant responded to 12 vignettes. Presentation of each vignette was followed by asking participants to rate the degree to which an agent's action was the cause of the outcome (*"To what extent was Carl's cleaning the cause behind the restaurant passing the inspection?"*). The ratings were made on an 11-point rating scale where participants could select whole

numbers ranging from 0 (“Not at all”) to 10 (“Completely”). On a separate page, participants also provided the blame or praise rating of the agent (*“How much blame or praise should Carl receive?”*). This rating was made on a common blame-praise scale that ranged from -5 (“Extreme Blame”) to +5 (“Extreme Praise”) with 0 denoting neither blame nor praise. Each vignette was accompanied by a check question (*“What did Carl do for the inspection?”*).

4. 2 Results

4.2.1 Causality Rating

A three way ANOVA of the causality ratings revealed a main effect of intention $F(1,41) = 19.50$, $p < .001$, $\eta_p^2 = .322$ and scenario $F(2,82) = 12.87$, $p < 0.05$, $\eta 2 = .100$. A significant interaction was recorded between intention and outcome $F(1, 41) = 118.75$, $p < .001$, $\eta_p^2 = .743$. The main effect of intention and the interaction between intention and outcome were significant within each scenario. Collapsing against scenarios, causal judgments replicated the asymmetry that had been observed for responsibility judgments (see Figure 4). Accordingly, agents with good intentions were held more causal for bad outcomes than agent with bad intentions were held for good outcomes.

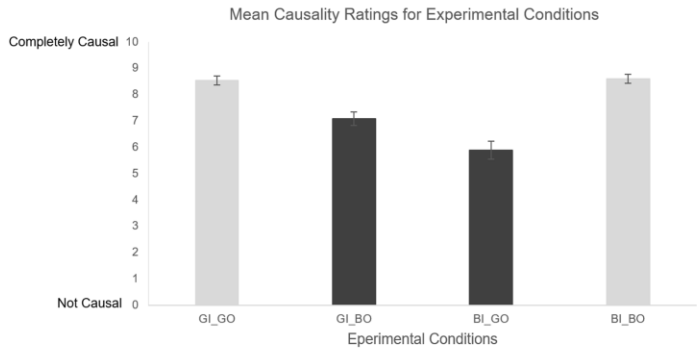

Figure 4. Mean causal ratings for the four conditions. Error bars indicate standard error of the mean. Higher scores reflect greater causal ratings. Abbreviation: G = Good; B = Bad; I = Intention; O = Outcome

4.2.2 Blame-Praise Ratings

A three way ANOVA for the blame-praise ratings revealed main effects of intention $F(1,41) = 147.62$, $p < .001$, $\eta_p^2 = .783$, and outcome $F(1,41) = 189.31$, $p < .001$, $\eta_p^2 = .822$, as well as a significant interaction between the two $F(1,41) = 11.54$, $p < 0.05$, $\eta_p^2 = .220$.

Participants' blame – praise ratings mimicked their ratings of causality and responsibility (see Figure 5). On average, participants choose to blame an agent when her good intentions led to negative outcomes (M = -1.25, SD = 1.46) but neither blame nor praise an agent when her bad intentions lead to positive outcomes (M = 0.73, SD = 1.61).

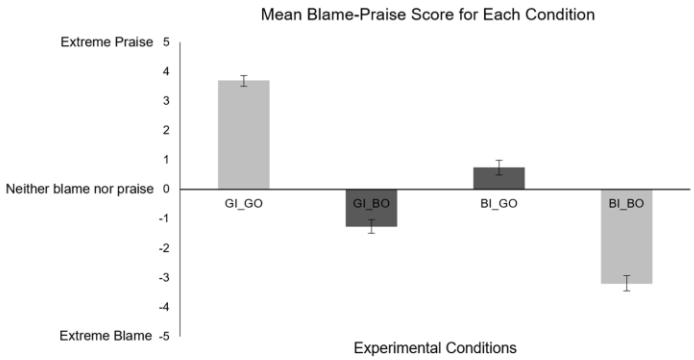

Figure 5. Mean blame-praise ratings. Higher scores reflect greater praise, lower negative scores reflect greater blame. A score of zero reflects neither blame nor praise.
Abbreviations: G = Good; B = Bad; I = Intention; O = Outcome

4.3 Discussion

The ratings of causality obtained in Experiments 3 mimic the ratings of responsibility obtained in the previous two experiments. This is intriguing. Predictions made by the two accounts diverge on expected judgments of causality. We would expect causal judgments to be high, for both incongruent agents, according to the hierarchical account and low according to the intentional account. Yet, instead of unequivocally supporting either perspective, our data presents evidence for an interaction between intention and outcome. Good intentioned agents with bad outcome are held more causal (just like they were held more responsible) than bad intentioned agents with good outcomes. Results from the experiment also assuage concerns regarding a confounding effect of the term responsibility. It appears that the observed asymmetry is not a result of the specific terminology or scenarios. Rather, the persistence of the asymmetry for causality ratings alludes to a difference in the evaluation of the two incongruent conditions. The exact reason for this is yet unknown. However, Experiment 4 explores one possible explanation.

Blame and praise ratings reaffirm the observed asymmetry. The experiment presents participants with a common blame-praise scale allowing them to choose between allocating

blame or praise on any aspect of the scenario. They can choose to focus on the intention of an agent, the outcome of the situation, neither of the two, or a combination of both. The results however reveal an interaction between intention and outcome such that average score inclined towards blame for the good intention-negative outcome agent and marginally towards praise for the bad intention-positive outcome agent. The difference in the degree of blame or praise allocated relative to baseline (which is neither praise nor blame) reflects the asymmetry between mismatched intentions and outcomes that has previously been observed for causality and responsibility judgments.

5. Experiment 4

The first three experiments present compelling evidence for an interaction between intention and outcome when they mismatch. Agents with good intentions are held more responsible, more causal, and more blameworthy for bad outcomes than are agents with bad intentions held for good outcomes. The sole objective of the fourth experiment is to explore the reasons for the asymmetrical judgments. While we can think of many different explanations, in this experiment we focus on one potential reason and leave consideration of alternatives to the general discussion. We propose that the asymmetry in the incongruent cases might be due to the participants making an additional causal inference about the agent's action and its impact on the overall outcome.

More specifically, a causally constructed chain typically has three components – identification of the mental representation of a desired end-state (intention), the means employed to bring about the outcome (action), and the outcome (Diks & Aarts, 2007). In the scenarios we present to participants, we systematically vary and explicitly provide information on two of the three components. People know that Sandra has a good or bad intention and the company loses or wins the contract. However, they know nothing about the action linking the intention with the outcome. In other words, participants have no explicit information regarding the how good

the presentation itself was – did Sandra make a great presentation? Or was her presentation terrible? Abundant research in social perception suggests that people often go beyond given behavioural information, including constructing social causal inferences (Heider & Simmel, 1944; Gilbert, 1989; Uleman, Newman, & Moskowitz, 1996) to enable them to understand behaviour better (Read, 1987). Previous research has shown that when presented with information on at least one component (from the three), people have the tendency to infer information on the other components automatically (Hassin, Bargh, & Uleman, 2002). We suspect that in the absence of information regarding the action component people might be inferring the state of the action in a way that justifies the outcome. In other words, we believe that in Sandra's case, the loss of the contract might be leading people to infer that she made a terrible presentation despite her good intentions. If this is indeed the case, making information regarding the action explicit should take away the asymmetry we have been observing. This is because the state of the action is directly under the control of the agent whereas the eventual outcome is not. If Sandra, in her desire to win the contract, made a great presentation but still lost the contract, she would arguably be held less responsible and less blameworthy because she did the best with the outcome directly under her control.

Experiment 4 systematically varied information regarding the state of the action, in addition to the intention and outcome, to assess its impact on subsequent judgements of causality and blame – praise. A sentence regarding the state of the action was added to previous scenarios. The action performed was either consistent with the intention or counter to the intention. For instance, when Sandra's intention was good (she wanted to obtain the contract) but her outcome was bad (she lost the contract) in the consistent-with-intention condition, she made a great presentation; in the counter-to-intention condition she made a terrible presentation. Note, her intention to get the contract and the outcome of losing the contract remained fixed. The only

information added was whether the action performed was consistent or counter to her intention. Similar variations were applied to the other cases.

5.1 Method

5.1.1 Participants

52 people participated in the experiment initially. 11 participants left the study prematurely, one participant failed the question checks, and another took more than three times the average completion time. After the elimination of these participants, the final sample of 39 was made up by 16 women (42.1%). The average age of a participant was 31.34 (*SD* = 11.28), range 18 to 60.

5.1.2 Design and Materials

The action factor, varied as consistent-with-intention or counter-to-intention, was added to the initial design of 2 x 2 x 3 (intention, outcome, scenario respectively). Consistent-with-intention was represented as a match between the agent's intention and the immediate outcome under her control, whereas counter-to-intention was presented as a mismatch between the agent's intention and the immediate outcome under her control (see Figure 6). All four factors were presented within-subject.

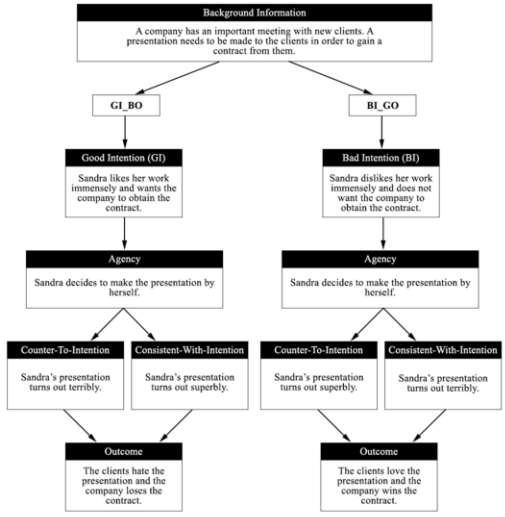

Figure 6. Parametric variations of the two action levels for the two incongruent conditions.

5.2 Result

5.2.1 Causal Ratings

According to the results of a four way repeated measures ANOVA, intention $F(1, 37) = 7.90$, $p < 0.01$, $\eta_p^2 = .176$ and scenario $F(2, 74) = 5.29$, $p < 0.01$, $\eta_p^2 = .125$ had main effects with good intentions ($M = 7.42$) scoring higher on average than bad intentions ($M = 7.06$). A significant interaction was also observed between intention and outcome $F(1, 37) = 39.97$, $p < .001$, $\eta_p^2 = .519$.

Action had no main effect $F(1, 37) = .70$, $p = .407$, ns on the data. However, as was expected, it did have a significant interaction with intention and outcome $F(1, 37) = 69.01$, $p < .001$, $\eta_p^2 = .651$ (see Figure 7). In line with our expectations, the asymmetry between the two incongruent cases did reach significance for the counter-to-intention condition, and failed to reach significance for the for the consistent-with-intention condition.

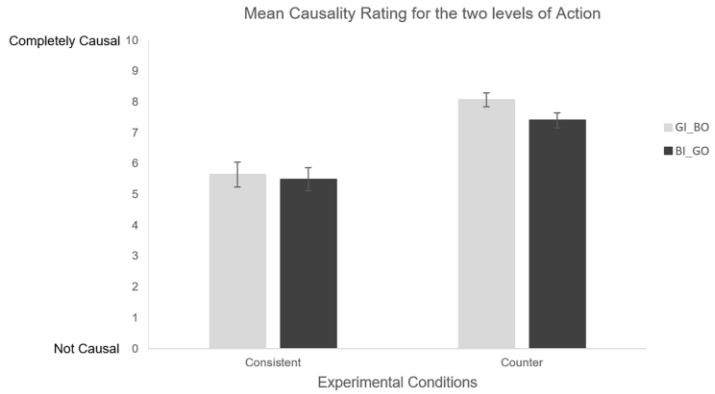

Figure 7. Mean causal ratings for the four conditions. Higher scores reflect greater causal ratings. Error bars represent standard error of the mean. Abbreviations: GI_BO = Good Intention Bad Outcome; BI_GO = Bad Intention Good Outcome

5.2.2 Blame-Praise Ratings

Mean blame-praise scores for all the four conditions further supports the suggestion about people's rich inferences (see Figure 8). The score awarded to both incongruent conditions under the consistent with intention condition is similar and practically zero (M = 0.34 for good intention-negative outcome and M = 0.29 for bad intention-positive outcome). However, in the counter to intention condition, allocation of blame-praise mimics that of Experiment 2 as good intention-negative outcome agent receive blame (M = -2.07) on average while bad intention-positive outcome agent receive marginal praise (M = .30).

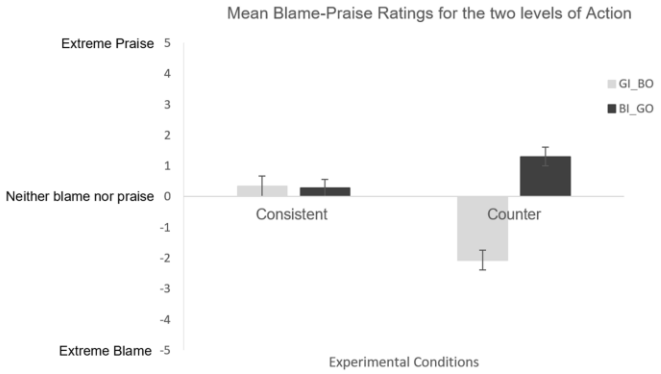

Figure 8. Mean blame-praise ratings. Higher scores reflect greater praise, lower negative scores reflect greater blame. A score of zero reflects neither blame nor praise. Abbreviations: GI_BO = Good Intention Bad Outcome; BI_GO = Bad Intention Good Outcome

5.3 Discussion

The aim of the present experiment was to examine one potential reason for the intention-outcome asymmetry between the two incongruent conditions. We hypothesised that participants were making inferences regarding the state of the action linking the intention and the outcome in a way that justified the attainment of the outcome. By systematically varying and explicitly stating information regarding an agent's action, we expected to see the asymmetry disappear under the consistent-with-intention condition and persist under the counter-to-intention condition. This was so because, the consistent-with-intention condition showed that an agent achieved the outcome under her control in accordance with her intention despite the eventual outcome, which was beyond her control, going in the opposite direction. In contrast, under the counter to intention condition, agent's actions ran counter to their intentions but match the outcome. The counter to intention condition thus presents the action in a way that justifies the attainment of the outcome and we expected to see the asymmetry under the counter to intention condition. Results from the study support our expectations. The asymmetry for both causal and blame – praise ratings was absent in the consistent with intention condition but persisted in the counter to intention condition. The results suggest that (among other explanations), participants

use inferences about the nature of the action to make their moral and causal judgments. Moreover, like previous experiments, neither the hierarchical nor the intentional perspective accounts for the pattern of data observed. Rather, the data demonstrate an interaction between intention and outcome (for a more detailed discussion on the topic, please see the general discussion).

6. General Discussion

The present set of experiments explored the impact of incongruence between intentions and outcomes on judgments of causality, responsibility, and blame and praise. We demonstrated two main findings. First, an asymmetry in the evaluation of cases when intentions mismatch with outcomes, according to which agents with good intentions and bad outcomes are held more responsible (Experiment 1 and Experiment 2), more causal (Experiment 3) and more blameworthy (Experiment 3) compared to agents with bad intentions and good outcomes. This finding cannot be explained on either the hierarchical account or intentional account of moral judgment. Second, in the presence of mismatched intentions and outcomes, participants draw inferences regarding the actions that link mental states to outcomes in a manner that justifies the outcome, thus producing the asymmetrical moral and causal judgments (Experiment 4).

6.1 Incongruence vs Congruence

In each of our experiments we note reductions in responsibility ratings, and blame and praise ratings for the two incongruent cases in comparison with the congruent cases. In other words, when an agent's intentions, whether good or bad, do not manifest into desired outcomes, the attributed moral accountability for the outcomes is reduced. This result finds substantial support from previous research as well as existing theoretical perspectives (Mikhail, 2007; Pillsbury, 2000; Cushman, Young, Hauser, 2006). Cushman (2008) reports an overall reduction in blame and wrongness judgments for conditions of accidental and attempted harm compared with congruent conditions depicting bad intentions manifesting into harmful outcomes.

Theoretical predictions from the hierarchical and the intentional accounts converge on attributing lenient moral judgments for unintended consequences. The overall reduction in responsibility and blame and praise ratings for the mismatched cases fits neatly with existing research.

Slightly less clear is the reduction observed in the causal ratings for the incongruent conditions compared to the congruent conditions. The predictions derived from the two theoretical perspectives diverge. On the hierarchical account we would expect to see both incongruent agents being held highly causal while on the intentional account we would expect to see a reduction in their causal association to the outcomes. This is because the hierarchical account subscribes to a hierarchical organisation of its factors with causal analysis preceding intentional analysis (Heider, 1958; Darley & Schulz, 1990; Shaver, 1985; Weiner, 1995). Therefore, a factual association between the agent and the outcome would be sufficient to regard high causality to the agent (even though the degree of responsibility might be reduced). On the intentional account however, causal and intentional analysis may influence one another simultaneously such that knowledge of the agent's intentions may alter the perception of causal association between the agent and the outcome. This stems from the postulation that moral norms may influence perceptions of causality (Alicke, 1992; Knobe & Fraser, 2008; Knobe, 2010; Kominsky et al., 2015). The overall reduction in causality ratings for the two incongruent condition reported in our experiments seems to support the intentional perspective, but this support is restricted as we do not explicitly test whether the reduction in causality ratings is due to the implicit influence of norms or if participant's are perceiving the causal relations in the conditions differently. Samland and Waldmann (2014) argue that findings showing altered causal judgments in morally relevant situations stem from ambiguity in the style of questioning rather than from the influence of moral norms. The majority of research on moral and causal judgments relies on vignettes to present participants with relevant information. Samland and

Waldmann (2016) propose that the arrangement of causal information alongside intentions and outcome information creates ambiguity that may lead people to interpret a question about causal judgment as a request to assess the agent's moral accountability instead of the causal relations. Our experiments employ vignettes and as such do not systematically untangle the learning of the causal relations from learning about the intentions and outcomes. Consequently, we cannot say for certain if the judgments we have obtained from participants' about causality reflect their perceptions on causal relations or are an expression of their judgment of moral accountability. While this clarification does not affect the larger picture that suggests differential evaluation of agents based on their intentions and outcomes, it will reveal the extent of the intention-outcome asymmetry effect. In other words, getting participants to answer questions about the causal relations in cases of intention – outcome mismatch would help identify if the incongruence affects only moral judgments or if it also distorts our perceptions of causation. Targeted research aimed at disentangling causal information from other moral information in moral and non-moral contexts will be a fruitful approach to understand the interplay between these factors.

6.2 Intention-Outcome Asymmetry Effect

A novel finding of the current work is an asymmetry in moral and causal judgments in response to the incongruence between intentions and outcomes. Agents with good intentions are held more morally and causally accountable for negative outcomes than agents with bad intentions are held for producing positive outcomes. This effect is peculiar given that in both conditions the agents are equally unsuccessful in bringing about their desired end-states. Hierarchical and intentional accounts would predict a reduction in moral judgments for cases of incongruence but no further nuanced difference between the two conditions of incongruence. Yet, our findings indicate a persistent asymmetry in evaluation of the two incongruent conditions.

In the final experiment we explored whether inference regarding the actions linking intentions with outcomes is producing the asymmetry. Research in social and personality psychology has shown that people often infer more information than has been provided, especially when the experimental stimuli use vignettes (Graesser, Singer, & Trabasso, 1994). We presented participants with the same set of incongruent stimuli as before but added a line regarding the nature of an agent's action. The action performed by an agent was either consistent with her intention or counter to it. The rationale for varying information regarding the action rested on the premise that participants were inferring the nature of the action to justify the outcomes. In our case it would mean that the asymmetry resulted from the participants inferring the action to be counter to the intentions (or consistent with the outcome). Results from the experiment supported our supposition. When the nature of the action variable made it explicit that the agent acted in agreement with their intention, despite the eventual outcome being contrary to her intention, participants reported no difference in the causal and blame judgments for the two incongruent agents. In other words, if an agent did the task under her control in agreement with their intention, they were considered to be less causal and responsible for the overall outcome even when the outcome was unintended, and this judgment was the same for agents who had a bad intention and brought about a positive outcome or those who had a good intention and brought about a negative outcome. However, when the information in the vignette revealed that the agent acted counter to her intentions but consistent with the outcome, the asymmetry not only reappeared, the overall degree of causality and blame attributed increased. We assume the re-appearance of the asymmetry as well as the overall increase stemmed from participants inferences being validated by the information provided.

It is important to note that while our account is supported by the experimental evidence, a number of alternative explanations exist. The final experiment provides participants with information regarding the consistency of an agent's action relative to her intention. However, it

could be that instead of the consistency, it is the attainment (or lack of) of the immediate outcome under the agent's control that affects the asymmetry. This perspective to understand the results of the final experiment is in accordance with our account, but it provides for a slightly different functional and mechanistic framework to understand the results.

An alternative explanation of the asymmetry is that different information may have different inherent value. There is some empirical support in favour of this assumption (Alicke et al., 2015; Uttich & Lombrozo, 2010). Both conditions of incongruence provide participants with two principle inputs – what an agent desired and what came of the situation. It could be the case that, specific combinations of intention and outcome have different inferential or communicative value. In other words, knowing of an agent who gets a negative outcome despite good intentions might implicitly communicate different information compared with that conveyed by knowing of an agent who has a bad intention but achieves a positive outcome, presumably about the agent's competence, effort, or character.

7. Conclusion

In a series of experiments we have identified a novel asymmetry in people's judgments of causality, responsibility, and blame. When intentions are incongruent with outcomes, people assign greater responsibility, greater causality, and greater blame to an agent with good intentions who produces a bad outcome than to an agent with bad intentions who produces a good outcome. We explored one possible explanation for this asymmetry, in terms of the additional inferences that people make beyond the information given in the scenarios, in order to make sense of the overall story. In particular, people seem to infer that a good intentioned agent who produces a bad outcome failed to perform the necessary action required to obtain the outcome and is thus more responsible than the bad intentioned agent who achieves a good outcome by chance. A key message from these findings is that in making responsibility and

causality judgments people invoke subtle extra inferences to make sense of incongruous patterns of events. Moral judgment and causal inference are closely intertwined.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63,* 368–378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126,* 556-574.

Arkkelin, D., Oakley, T., & Mynatt, C. R. (1979). Effects of controllable versus uncontrollable factors on responsibility attributions: A single-subject approach. *Journal of Personality and Social Psychology, 37,* 110-115.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94,* 74–85.

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass, 8,* 536-554.

Coffman, L.C. (2011). Intermediation reduces punishment (and reward). *American Economic Journal: Microeconomics,* 77-106.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108,* 353-380.

Cushman, F., Skeketoff, R., & Wharton, S., & Carey, S. (2013). The development of inten-based moral jugdment. *Cognition, 127,* 6-21.

Cushman, F., Young, L., & Hauser, M. D. (2006). The role of conscious reasoning and intuitions in moral judgment: Testing three principles of harm. *Psychological Science, 17,* 1082-1089.

Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual review of psychology, 41,* 525-556.

DeJoy, D. M., & Klippel, J. A. (1984). Attributing responsibility for alcohol-related near-miss accidents. *Journal of Safety Research, 15,* 107-115.

Dik, G., & Aarts, H. (2007). Behavioral cues to others' motivation and goal pursuits: The perception of effort facilitates goal inference and contagion. *Journal of Experimental Social Psychology, 43*, 727-737.

Fincham, F. D., & Jaspars, J. M. (1980). Attribution of responsibility: From man the scientist to man as lawyer. In L. Berkowitz (Ed.), *Advances in experimental social psychology, 13*, 81-138. New York: Academic Press.

Fincham, F. D., & Roberts, C. (1985). Intervening causation and the mitigating responsibility for harm doing. *Journal of Experimental Social Psychology, 21*, 178-194.

Gerstenberg, T., Lagnado, D. A., & Kareev, Y. (2010). The dice are cast: The role of intended versus actual contributions in responsibility attribution. In S. Ohlsson & R. Catrambone (Eds.), *Cognition in flux: Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, (pp. 1697-1702). Austin, TX: Cognitive Science Society.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189-211). NewYork: Guilford Press.

Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgments. *Harvard Business School NOM Working Paper*, 08-80.

Graesser, C. A., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371-395.

Guglielmo, S. (2015). Moral judgment as information processing: an integrative review. *Frontiers in Psychology, 6*, 1637.

Hassin, R. R., Bargh, J. A., & Uleman, J. S. (2002). Spontaneous causal inferences. *Journal of Experimental Social Psychology, 38*, 515–522.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243-259.

Hitchcock, C. & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 11*, 587-612.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*, 315-329.

Knobe, J. & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnot-Armstrong (Ed.), *Moral Psychology*, vol. 2: *The Cognitive Science of Morality: Intuition and Diversity.* Cambridge, MA: MIT Press.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D. A., & Knobe, J. (2015). Causal Superseding. *Cognition, 137*, 196-209.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition, 108*, 754-770.

Lombrozo, T. (2010). Causal-explanatory pluralism: how intention, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology, 61*, 303-332.

Lowe, C. A., & Medway, F. J. (1976). Effects of valence, severity, and relevance on responsibility and dispositional attribution. *Journal of Personality, 44*, 518-538.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*, 147-186.

Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology, 26*, 131-146.

McClure, J., Hilton, D.J., & Sutton, R. M. (2007). Judgments of voluntary and physcial causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European Journal of Social Psychology, 37*, 879-901.

McMartin, J. A., & Shaw, J. I. (1977). An attributional analysis of responsibility for a happy accident: Effects of ability, intention, and effort. *Human Relations, 30*, 899-918.

Medway, F. J., & Lowe, C. A. (1975). Effects of outcome valence and severity on attribution of responsibility. *Psychological Reports, 36,* 239-246.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in cognitive sciences, 11,* 143-152.

Phares, E. J., & Wilson, K. G. (1972). Responsibility attribution: Role of outcome severity, situational ambiguity, and internal-external control. *Journal of Personality, 40,* 392-406.

Phillips, J. & Shaw, A. (2014). Responsibility attribution: Role of outcome severity, situational ambiguity, and internal-external control. *Journal of Personality, 40,* 392-406.

Pillsbury, S. H. (2000). *Judging evil: Rethinking the law of murder and manslaughter.* NYU Press.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology, 39,* 653-660.

Provencher, H. L., & Fincham, F. D. (2000). Attributions of causality, responsibility and blame for positive and negative symptom behaviours in caregivers of persons with schizophrenia. *Psychological Medicine, 30,* 899-910.

Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology, 52,* 288–302.

Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences, *Cognition, 156,* 164-176.

Shaver, K. (1985). *TheAttribution of blame : Causality, responsibility, and blameworthiness.* New York, NY: Springer.

Shaver, K. G. (1970). Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology, 14,* 101-103.

Shaw, J. I., & McMartin, J. A. (1977). Personal and situational determinants of attribution of responsibility for an accident. *Human Relations, 30,* 95-107.

Shaw, J. I., & Skolnick, P. (1971). Attribution of responsibility for a happy accident. *Journal of Personality and Social Psychology, 18,* 380-383.

Shultz, T. R., Schleifer, M., & Altman, I. (1981). Judgments of causation, responsibility, and punishment in cases of harm-doing. *Canadian Journal of Behavioural Science, 13*, 238-253.

Slain, A. J., Penrod, S., Garbin, C. P., & Stolle, D. P. (1998). Multidimensional perceptions of injury: Implications for the "adversary culture". In AP-LS Conference, Redondo Beach, CA.

Thomas, E. A., & Parpal, M. (1987). Liability as a function of plaintiff and defendant fault. *Journal of Personality and Social Psychology, 53,* 843-857.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inferences. In M. P. Zanna (Ed.), *Advances in experimental social psychology (Vol. 28*, pp. 211–279). New York: Academic Press.

Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition, 116,* 87–100.

Walster, E. (1966). Assignment of responsibility for an accident. *Journal of personality and social psychology, 3,* 73-79.

Walster, E. (1967). "Second guessing" important events. *Human Relations, 20,* 239-250.

Weiner, B. (1995). *Judgments of responsibility : A foundation for a theory of social conduct.* New York: Guilford Press.

Wissler, R., Evans, D. L., Hart, A. J., Morry, M. M., & Saks, M. J. (1997). Explaining "pain and suffering" awards: The role of injury characteristics and fault attributions. *Law and Human Behavior, 21,* 181–207.

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120,* 202-214.

Appendix A

Table A1

Company presentation scenario

		Outcome	
		Positive	Negative
Intention	Good	*Sandra likes her work immensely and wants the company to obtain the contract. Sandra decides to make the presentation by herself. The clients love the presentation and the company wins the contract.*	*John likes his work immensely and wants the company to obtain the contract. John decides to make the presentation by himself. However, the clients hate the presentation and the company loses the contract.*
	Bad	*Anna dislikes her work immensely and does not want the company to obtain the contract. Anna decides to make the presentation by herself. However, the clients love the presentation and the company wins the contract.*	*Mark dislikes his work immensely and does not want the company to obtain the contract. Mark decides to make the presentation by himself. The clients hate the presentation and the company loses the contract.*

Note. Constant background information: *A company has an important meeting with new clients. A presentation needs to be made to the clients in order to gain a contract from them.*

Table A2

Restaurant cleaning inspection scenario

		Outcome	
		Positive	Negative
Intention	Good	*Carl finds his employment at the restaurant rewarding and wants the restaurant to pass the inspection. Carl volunteers to do the cleaning alone. The cleaning inspectors find the restaurant clean and the restaurant passes the examination.*	*Rosie finds her employment at the restaurant rewarding and wants the restaurant to pass the inspection. Rosie volunteers to do the cleaning alone. However, the cleaning inspectors find the restaurant dirty and the restaurant fails the examination.*
	Bad	*David finds his employment at the restaurant unrewarding and wants the restaurant to fail the inspection. David volunteers to do the cleaning alone. However, the cleaning inspectors find the restaurant clean and the restaurant passes the examination.*	*Tracy finds her employment at the restaurant unrewarding and wants the restaurant to fail the inspection. Tracy volunteers to do the cleaning alone. The cleaning inspectors find the restaurant dirty and the restaurant fails the examination.*

Note. Constant background information: *A restaurant has a cleaning inspection coming up. The restaurant needs to pass the inspection in order to maintain its standard of health and hygiene*

Table A3

Theatre stage production scenario

		Outcome	
		Positive	Negative
Intention	Good	*Greg gets along with the management team and wants to increase the theatre company's popularity. Greg volunteers to direct the production on his own. The audiences enjoy the production and the theatre company earns a high reputation.*	*Sophie gets along with the management team and wants to increase the theatre company's popularity. Sophie volunteers to direct the production on her own. However, the audiences are bored with the production and the theatre company earns a low reputation.*
	Bad	*Peter does not get along with the management team and wants to decrease the theatre company's popularity. Peter volunteers to direct the production on his own. However, the audiences enjoy the production and the theatre company earns a high reputation.*	*Isabella does not get along with the management team and wants to decrease the theatre company's popularity. Isabella volunteers to direct the production on her own. The audiences are bored with the production and the theatre company earns a low reputation.*

Note. Constant background information: *A theatre company is preparing a stage production. The production is an opportunity for the theatre company to display their work to enhance their popularity.*

Table A4

Family gathering scenario

		Outcome	
		Positive	Negative
Intention	Good	*Emily adores the bride and wants the bride and the guests to have an enjoyable family gathering. Emily decides to organize the entire event on her own. The bride and the relatives love the arrangements and the gathering is a huge success.*	*Andrew adores the bride and wants the bride and the guests to have an enjoyable family gathering. Andrew decides to organize the entire event on his own. However, the bride and the relatives hate the arrangements and the gathering is a huge failure.*
	Bad	*Jennifer detests the bride and wants the bride and the guests to have a terrible family gathering. Jennifer decides to organize the entire event on her own. However, the bride and the relatives love the arrangements and the gathering is a huge success.*	*Brent detests the bride and wants the bride and the guests to have a terrible family gathering. Brent decides to organize the entire event on his own. The bride and the relatives hate the arrangements and the gathering is a huge failure.*

Note. Constant background information: *A couple is getting married and wedding festivities are being planned. A family gathering needs to be organized in order for the relatives and the couple to relax before the wedding.*

Appendix B

Table B1
Gardening scenario (low intensity)

		Outcome	
		Positive	Negative
Intention	Good	*Their son, Thomas, shares their passion for gardening and wants the backyard to be converted into a garden. Thomas decides to do the gardening on his own. Within a short period of time, the ground becomes fertile and small plants appear.*	*Their daughter, Sarah, shares their passion for gardening and wants the backyard to be converted into a garden. Sarah decides to do the gardening on her own. However within a short period of time, the ground becomes infertile and no plants appear.*
	Bad	*Their son, Alex, does not share their passion for gardening and does not want the backyard to be converted into a garden. Alex decides to do the gardening on his own. However within a short period of time, the ground becomes fertile and small plants appear.*	*Their daughter, Patricia, does not share their passion for gardening and does not want the backyard to be converted into a garden. Patricia decides to do the gardening on her own. Within a short period of time, the ground becomes infertile and no plants appear.*

Note. Constant background information: *An elderly couple owns a house with a backyard. The couple wants to convert the backyard into a garden.*

Table B2
Gardening scenario (high intensity)

		Outcome	
		Positive	Negative
Intention	Good	*Their son, James, shares their passion for gardening and wants the backyard to be converted into a garden. James decides to do the gardening on his own. Within a short period of time, the ground becomes fertile. The backyard turns into a beautiful garden and the garden becomes a public attraction for the entire town.*	*Their daughter, Mary, shares their passion for gardening and wants the backyard to be converted into a garden. Mary decides to do the gardening on her own. However within a short period of time, the ground becomes infertile. The backyard is completely ruined and it becomes an eyesore for the entire town.*
	Bad	*Their son, Robert, does not share their passion for gardening and does not want the backyard to be converted into a garden. Robert decides to do the gardening on his own. However within a short period of time, the ground becomes fertile. The backyard turns into a beautiful garden and the garden becomes a public attraction for the entire town.*	*Their daughter, Linda, does not share their passion for gardening and does not want the backyard to be converted into a garden. Linda decides to do the gardening on her own. Within a short period of time, the ground becomes infertile. The backyard is completely ruined and it becomes an eyesore for the entire town.*

Note. Constant background information: *An elderly couple owns a house with a backyard. The couple wants to convert the backyard into a garden*

Table B3
School prom party (low intensity)

		Outcome	
		Positive	Negative
Intention	Good	*Susan adores her sister and wants the sister to look good at the party. Susan decides to arrange her sister's dress alone. The dress fits well and the sister is happy with the way it looks on her.*	*Michael adores his sister and wants the sister to look good at the party. Michael decides to arrange his sister's dress alone. However the dress fits badly and the sister is unhappy with the way it looks on her.*
	Bad	*Lisa detests her sister and wants the sister to look bad at the party. Lisa decides to arrange her sister's dress alone. However the dress fits well and the sister is happy with the way it looks on her.*	*William detests his sister and wants the sister to look bad at the party. William decides to arrange his sister's dress alone. The dress fits badly and the sister is unhappy with the way it looks on her.*

Note. Constant background information: *A school year is coming to an end and a prom party has been organized. All of the girls are excited about the dresses that they will wear.*

Table B4
School prom party (high intensity)

		Outcome	
		Positive	Negative
Intention	Good	*Nancy adores her sister and wants the sister to look good at the party. Nancy decides to arrange her sister's dress alone. The dress fits perfectly. The sister is the best dressed and is chosen as the Prom Queen.*	*David adores his sister and wants the sister to look good at the party. David decides to arrange his sister's dress alone. However the dress fits horribly. The sister is the worst dressed and gets bullied at school.*
	Bad	*Helen detests her sister wants the sister to look bad at the party. Helen decides to arrange her sister's dress alone. However the dress fits perfectly. The sister is the best dressed and is chosen as the Prom Queen.*	*John detests his sister wants the sister to look bad at the party. John decides to arrange his sister's dress alone. The dress fits horribly. The sister is the worst dressed and gets bullied at school.*

Note. Constant background information: *A school year is coming to an end and a prom party has been organized. All of the girls are excited about the dresses that they will wear.*

Table B5

House redecoration (low intensity)

		Outcome	
		Positive	Negative
Intention	Good	*Their friend, Charles, likes the couple and wants the redecoration to go well. Charles decides to redecorate the house alone. The redecoration is completed within the deadline and the house looks good.*	*Their friend, Carol, likes the couple and wants the redecoration to go well. Carol decides to redecorate the house alone. However, the redecoration is not completed within the deadline and the house looks bad.*
	Bad	*Their friend, Joseph, does not like the couple and does not want the redecoration to go well. Joseph decides to redecorate the house alone. However, the redecoration is completed within the deadline and the house looks good.*	*Their friend, Amy, does not like the couple and does not want the redecoration to go well. Amy decides to redecorate the house alone. The redecoration is not completed within the deadline and the house looks bad.*

Note. Constant background information: *A husband and wife recently bought a new home. The house needs to be redecorated within a deadline so that the couple can move in.*

Table B6

House redecoration (high intensity)

		Outcome	
		Positive	Negative
Intention	Good	*Their friend, Steve, likes the couple and wants the redecoration to go well. Steve decides to redecorate the house alone. The redecoration is completed within the deadline. The house looks beautiful and its value doubles in price.*	*Their friend, Amanda, likes the couple and wants the redecoration to go well. Amanda decides to redecorate the house alone. However, the redecoration is not completed within the deadline. The house looks ugly and its value halves in price.*
	Bad	*Their friend, Mark, does not like the couple and does not want the redecoration to go well. Mark decides to redecorate the house alone. However, the redecoration is completed within the deadline. The house looks beautiful and its value doubles in price.*	*Their friend, Kate, does not like the couple and does not want the redecoration to go well. Kate decides to redecorate the house alone. The redecoration is not completed within the deadline. The house looks ugly and its value halves in price.*

Note. Constant background information: *A husband and wife recently bought a new home. The house needs to be redecorated within a deadline so that the couple can move in.*

List of figures with figure captions